

Document made available under the Patent Cooperation Treaty (PCT)

International application number: PCT/JP05/006397

International filing date: 31 March 2005 (31.03.2005)

Document type: Certified copy of priority document

Document details: Country/Office: JP
Number: 2004-108456
Filing date: 31 March 2004 (31.03.2004)

Date of receipt at the International Bureau: 09 June 2005 (09.06.2005)

Remark: Priority document submitted or transmitted to the International Bureau in compliance with Rule 17.1(a) or (b)



World Intellectual Property Organization (WIPO) - Geneva, Switzerland
Organisation Mondiale de la Propriété Intellectuelle (OMPI) - Genève, Suisse

日 本 国 特 許 庁
JAPAN PATENT OFFICE

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office.

出 願 年 月 日
Date of Application: 2 0 0 4 年 3 月 3 1 日

出 願 番 号
Application Number: 特 願 2 0 0 4 - 1 0 8 4 5 6

パリ条約による外国への出願
に用いる優先権の主張の基礎
となる出願の国コードと出願
番号
J P 2 0 0 4 - 1 0 8 4 5 6
The country code and number
of your priority application,
to be used for filing abroad
under the Paris Convention, is

出 願 人
Applicant(s): 株式会社バイオシンクタンク

2 0 0 5 年 5 月 2 5 日

特許庁長官
Commissioner,
Japan Patent Office

小 川



【書類名】	特許願
【整理番号】	BIOTT-0002
【あて先】	特許庁長官 殿
【国際特許分類】	G06F 17/30
【発明者】	
【住所又は居所】	東京都練馬区田柄 1 丁目 1 9－7
【氏名】	森下 真一
【発明者】	
【住所又は居所】	東京都文京区本郷 5 丁目 1 4－8 テラス本郷 3 0 2 号室
【氏名】	山田 智之
【特許出願人】	
【住所又は居所】	東京都練馬区田柄 1 丁目 1 9－7
【氏名又は名称】	森下 真一
【特許出願人】	
【住所又は居所】	東京都文京区本郷 5 丁目 1 4－8 テラス本郷 3 0 2 号室
【氏名又は名称】	山田 智之
【特許出願人】	
【住所又は居所】	神奈川県横浜市西区みなとみらい四丁目 1 0 番 1－E 1 7 0 6 号
【氏名又は名称】	名取 幸和
【代理人】	
【識別番号】	100109553
【弁理士】	
【氏名又は名称】	工藤 一郎
【手数料の表示】	
【予納台帳番号】	100322
【納付金額】	21,000円
【提出物件の目録】	
【物件名】	特許請求の範囲 1
【物件名】	明細書 1
【物件名】	図面 1
【物件名】	要約書 1

【書類名】 特許請求の範囲

【請求項 1】

遺伝子情報を表わす遺伝子塩基配列を格納したデータベースを検索するための索引であり、所定の長さである所定長の塩基配列が前記遺伝子塩基配列の中に出現する位置を検索するための索引、を用いて、入力される塩基配列と同じ長さで類似する塩基配列であり前記遺伝子塩基配列に出現する塩基配列である類似塩基配列を検索するための塩基配列検索装置であって、

前記所定長を超える長さの塩基配列を入力する塩基配列入力部と、

前記塩基配列入力部に入力された塩基配列である入力塩基配列に対して、適合しない塩基への置換の操作を行なう塩基の個数を示すハミング距離を入力するハミング距離入力部と、

前記入力塩基配列の部分配列であって、前記所定長の長さを持ち異なる 2 つの部分配列と、その余の部分と、を特定する特定部と、

前記特定部で特定された部分配列とその余の部分とに、前記ハミング距離入力部で入力されたハミング距離を分割して割り当てる割当部と、

前記特定部で特定された 2 つの部分配列のうち、前記割当部で割り当てられたハミング距離で示される個数の塩基を適合しない塩基へ置換する操作を前記部分配列に対して行なって生成される塩基配列である置換塩基配列の総数が大きくない方を選択する選択部と、

前記選択部により選択された部分配列に対して、前記割当部で割り当てられたハミング距離をもつ置換塩基配列を生成する置換塩基配列生成部と、

前記置換塩基配列生成部で生成された部分配列を検索キーとして前記索引を用いて検索を行なう検索部と、

を有する塩基配列検索装置。

【請求項 2】

前記特定部は、

前記塩基配列入力部で入力された塩基配列の塩基数が前記所定長の 2 倍以下または 2 倍未満であれば、前記 2 つの部分配列のうち的一方の部分配列の端を前記入力塩基配列の一方の端と一致させ、前記 2 つの部分配列のうちの他方の部分配列の端を前記入力塩基配列の他方の端と一致させ、その余の部分が生じず特定されないことにする第一特定手段を有する請求項 1 に記載の塩基配列検索装置。

【請求項 3】

前記特定部は、

前記塩基配列入力部で入力された塩基配列の塩基数が前記所定長の 2 倍より大であれば、前記 2 つの部分配列が重ならないことにして前記 2 つの部分配列を特定する第二特定手段を有する請求項 1 または 2 に記載の塩基配列検索装置。

【請求項 4】

前記検索部での検索結果に基づいて、前記置換塩基配列を含んで遺伝子塩基配列に現れる塩基配列である類似候補塩基配列を取得する類似候補塩基配列取得部と、

前記類似候補塩基配列取得部で取得された類似候補塩基配列と前記入力塩基配列とのハミング距離が前記ハミング距離入力部に入力されたハミング距離と等しいかどうかを判定する判定部と、

を有する請求項 1 から 3 のいずれかーに記載の塩基配列検索装置。

【請求項 5】

適合しない塩基の組を指定する不適合塩基組入力部を有し、不適合塩基組入力部に入力された塩基の組に基づいて、検索部で検索が行なわれ、また、ハミング距離が求められる請求項 4 に記載の塩基配列検索装置。

【請求項 6】

前記塩基配列入力部に入力された塩基配列と類似塩基配列との対応する塩基の適合の分布を表わす分布情報を入力する適合分布入力部を有し、

前記判定部は、前記適合分布入力部で入力された分布情報が満たされているかどうかを

判定する分布判定手段を有する請求項 4 または 5 のいずれかに記載の塩基配列検索装置。

【請求項 7】

前記適合分布入力部で入力される分布情報は、塩基配列と類似塩基配列との対応する塩基が連続して適合する長さの下限である請求項 6 に記載の塩基配列検索装置。

【請求項 8】

前記塩基配列入力部に入力される塩基配列の長さは 15 から 25 であり、前記所定長は、11 から 14 である請求項 1 から 7 のいずれかに記載の塩基配列検索装置。

【請求項 9】

遺伝子情報を表わす遺伝子塩基配列を格納したデータベースを検索するための索引であって、所定の長さである所定長の塩基配列が前記遺伝子塩基配列の中に出現する位置を検索するための索引、を用いて、入力される塩基配列と同じ長さで類似する塩基配列であり前記遺伝子塩基配列に出現する塩基配列である類似塩基配列を検索するための塩基配列検索方法であって、

前記所定長を超えるの長さの塩基配列を入力する塩基配列入力ステップと、

前記塩基配列入力部に入力された塩基配列である入力塩基配列に対して、適合しない塩基への置換の操作を行なう塩基の個数を示すハミング距離を入力するハミング距離入力ステップと、

前記入力塩基配列の部分配列であって、前記所定長の長さを持ち異なる 2 つの部分配列と、その余の部分と、を特定する特定ステップと、

前記特定ステップで特定された 2 つの部分配列とその余の部分とに、前記ハミング距離入力ステップにて入力されたハミング距離を分割して割り当てる割当ステップと、

前記特定ステップで特定された 2 つの部分配列のうち、前記割当部で割り当てられたハミング距離で示される個数の塩基を適合しない塩基へ置換する操作を前記部分配列に対して行なって生成される塩基配列である置換塩基配列の総数が大きくない方を選択する選択ステップと、

前記選択ステップにより選択された部分配列に対して、前記割当ステップにて割り当てられたハミング距離をもつ置換塩基配列を生成する置換塩基配列生成ステップと、

前記置換塩基配列生成ステップで生成された部分配列を検索キーとして前記索引を用いて検索を行なう検索ステップと、

を含む塩基配列検索方法。

【請求項 10】

アルファベットが一次元に配列した文字列を格納したデータベースを検索するための索引であり、所定の長さである所定長の文字列が前記データベースに格納された文字列の中に出現する位置を検索するための索引、を用いて、入力される文字列と同じ長さで類似する文字列であり前記前記データベースに格納された文字列に出現する文字列である類似文字列を検索するための文字列検索装置であって、

前記所定長を超える長さの文字列を入力する文字列入力部と、

前記文字列入力部に入力された文字列である入力文字列に対して、適合しないアルファベットへの置換の操作を行なうアルファベットの個数を示すハミング距離を入力するハミング距離入力部と、

前記入力文字列の部分文字列であって、前記所定長の長さを持ち異なる 2 つの部分文字列と、その余の部分と、を特定する特定部と、

前記特定部で特定された部分文字列とその余の部分とに、前記ハミング距離入力部で入力されたハミング距離を分割して割り当てる割当部と、

前記特定部で特定された 2 つの部分文字列のうち、前記割当部で割り当てられたハミング距離で示される個数のアルファベットを適合しないアルファベットへ置換する操作を前記部分文字列に対して行なって生成される文字列である置換文字列の総数が大きくない方を選択する選択部と、

前記選択部により選択された部分文字列に対して、前記割当部で割り当てられたハミン

グ距離をもつ置換文字列を生成する置換文字列生成部と、

前記置換文字列生成部で生成された置換文字列を検索キーとして前記索引を用いて検索を行なう検索部と、

を有する文字列検索装置。

【請求項 11】

前記文字列は、ペプチド配列である請求項 10 に記載の文字列検索装置。

【書類名】 明細書

【発明の名称】 塩基配列検索装置及び塩基配列検索方法

【技術分野】

【0001】

本発明は、遺伝子情報を表わす遺伝子塩基配列を検索する装置及び方法に関する。

【背景技術】

【0002】

ワトソンとクリックとによるDNA (Deoxyribo Nucleic Acid) の構造の解明に基づき、塩基配列に基づく遺伝子情報の研究が発展している。DNAは、アデニン (A)、シトシン (C)、グアニン (G)、チミン (T) の塩基のいずれかを含むヌクレオチドが並んでいる構造を持ち、細胞の核の中では、通常、AとT、GとC、の結合により、二重らせんの構造となっている。遺伝子を表現するDNAのヌクレオチドの配列 (以下、「遺伝子塩基配列」と呼ぶ) が、RNA (Ribonucleic Acid) に転写され、スプライシングを経て、mRNA (messenger RNA) が生成され、たんぱく質の合成がされることが知られている。RNAは、D-リボースを糖成分として、アデニン (A)、シトシン (C)、グアニン (G)、ウラシル (U) を塩基とする核酸である。

【0003】

ところで、近年、RNA干渉と呼ばれる現象が発生することが知られるようになった。RNA干渉とは、細胞内に特定の2本鎖RNAが存在することにより、特定の配列のmRNAが破壊され、遺伝子の発現が抑制される現象である。この現象は、最初、線虫の細胞を用いた実験で見出された。その後、この現象は、哺乳動物細胞でも起きることが知られるようになり、注目を集めることとなった。人為的にRNA干渉を起こすことにより、特定の遺伝子の働きを抑制することができ、その特定の遺伝子の働きを調べることができるからである。また、RNA干渉を利用することにより、特定の遺伝子の働きを抑制する効果を発揮する薬を開発できる可能性も生まれてきた。

【0004】

図1は、RNA干渉の過程の概略を示す図である。RNA干渉は、以下のようなプロセスを経て発生すると考えられている。およそ21から23塩基対の長さのsiRNA (short interfering RNA) がマルチ・タンパク質複合体と結合し、RISC (RNA-induced silencing complex) を形成する。RISCは、そのsiRNAと相同性を持つmRNAと結合し、そのmRNAを分解することにより、そのmRNAが機能しなくなる。ここで、「相同性」とは、相補性が不完全な場合であり、「相補性」とは、二つの塩基配列の全体において、AとT、GとC、AとUとの対が形成されていることをいう。したがって、相同性とは、二つの塩基配列の一部に、AとT、GとC、AとU以外の対が発生していることを意味する。なお、これらの対以外に、GとUとの対が形成される可能性もあることが知られている。

【0005】

したがって、RNA干渉を発生させ、目的とする遺伝子の働きを抑制するためには、siRNAの配列を設計することが重要である。すなわち、目的とする遺伝子だけに現れ、他の遺伝子の塩基配列と相同性を持たない、siRNAの配列を設計することが重要である。したがって、siRNAの配列を設計する際には、siRNAの配列に似た塩基配列を持つ遺伝子が存在しないことを確認することが必要となる。

【0006】

また、近年、マイクロアレイを用いた遺伝子解析や遺伝子診断などが実施されている。「マイクロアレイ」とは、長さが15から30塩基程度のオリゴDNAをガラスなどの基板上に合成したDNAチップの一種である (例えば、非特許文献1参照。)。

【0007】

図2は、マイクロアレイを用いた遺伝子解析や遺伝子診断などの過程を例示する。ガラスなどの基板上に合成したオリゴDNAを持つマイクロアレイ201上に、蛍光色素などの標識203を付加されたDNA (202) を流すと、そのDNAと相補性あるいは相同

性を持つマイクロアレイ上のオリゴDNAとが結合（ハイブリダイズ）する（符号204）。どの場所のオリゴDNAとハイブリダイズしたかを、標識の蛍光色素による蛍光を検出することにより、DNA（202）の種類などを判定する。図2では、マイクロアレイ上に数本のオリゴDNAしか示されていないが、実際のマイクロアレイは、縦横の長さが0.5インチ程度の領域に万のオーダーでオリゴDNAが配置される。

【0008】

したがって、どのような塩基配列を持つオリゴDNAをマイクロアレイに配置するかを決めることは、マイクロアレイの設計において、極めて重要な工程である。

【0009】

従来においては、似た塩基配列が存在するかどうかの検出は、BLAST（例えば非特許文献2参照。）と呼ばれるソフトウェアや、Smith-Watermanと呼ばれるアルゴリズム（例えば、非特許文献3参照。）を用いた、遺伝子情報を表わす遺伝子塩基配列を格納したデータベースの検索により行なわれている場合が多い。

【非特許文献1】杉本直己著、“遺伝子化学”、19ページ、株式会社化学同人発行、2002年

【非特許文献2】S. F. Altschul, W. Gish, W. Miller, E. W. Myers, and D. J. Lipman, “Basic local alignment search tool”, J. Mol. Biol., 215, 403-410, 1990

【非特許文献3】T. F. Smith, and M. S. Waterman, “Identification of common molecular subsequences”, J. Mol. Biol., 147, 195-197, 1981

【発明の開示】

【発明が解決しようとする課題】

【0010】

しかしながら、BLASTを使用する方法では、似た塩基配列の存在を見落としてしまうという課題がある。BLASTは、通常、7塩基が連続して同じになる部分を用いて検索が行なわれる。このため、19塩基の塩基配列が与えられた場合、例えば、図3のXの位置で塩基の不一致あるいは不適合がある塩基配列を見つけることができず、似た塩基配列の存在を見落としてしまう。

【0011】

また、Smith-Watermanアルゴリズムを用いれば、似た塩基配列の存在を正しく検出することが可能であるが、必要とされる計算量が大きく、検出までに時間がかかるという課題がある。

【0012】

そこで、本発明は、似た塩基配列の存在を少ない計算量で検出することができる装置及び方法などを提供することを目的とする。

【課題を解決するための手段】

【0013】

かかる目的を達成するために、本発明では、入力された塩基配列から所定の長さの二つの部分配列とその余の部分とを特定して、対応する塩基が適合しない数であるハミング距離を、それらの部分配列とその余の部分とに分割して割り当てて、2つの部分配列のうち、それぞれの部分配列に対して、割り当てられたハミング距離の個数の塩基を適合しない塩基に置換する操作を加えて得られる塩基配列の総数が大きくないほうを選択して、検索を行なうようにする。

【0014】

これにより、検索に用いる塩基配列であって塩基を置換して生成する塩基配列の数を少なくすることができ、検索のために必要となる計算量を小さくすることができ、課題が解決される。また、ハミング距離が所定の値となる、似た塩基配列の存在を見落とすこともなくなるので、課題が解決される。

【発明の効果】

【0015】

本発明に係る塩基配列検索装置及び塩基配列検索方法によれば、検索のために必要となる計算量を小さくすることができ、また、ハミング距離が所定の値となり似た塩基配列の存在を見落とすことも無い。

【発明を実施するための最良の形態】

【0016】

以下、本発明を実施するための最良の形態について、図を用いて実施形態として説明する。なお、本発明は、これら実施の形態に何ら限定されるものではなく、その要旨を逸脱しない範囲において、種々なる態様で実施し得る。

【0017】

（実施形態1：主に請求項1、9について説明する）

【0018】

本発明の実施形態1として、遺伝子情報を表わす遺伝子塩基配列を格納したデータベースにおける所定長の塩基配列の出現を検索するための索引を用いて、類似塩基配列を検索するための塩基配列検索装置として、入力された塩基配列から所定の長さの二つの部分配列とその余の部分とを特定し、対応する塩基が適合しない数であるハミング距離を、それらの部分配列とその余の部分とに分割して割り当てて、2つの部分配列のうち、それぞれの部分配列に対して、割り当てられたハミング距離の個数の塩基を適合しない塩基に置換する操作を加えて得られる塩基配列の総数が大きくないほうを選択して検索を行なう塩基配列検索装置について説明する。

【0019】

なお、ここに、「所定長」とは所定の長さである。この所定の長さは、遺伝子塩基配列を格納したデータベースの検索を行なうための索引が受け付けることができる塩基配列の長さである。例えば、BLASTの場合には、通常は、所定長は7となる。また、「類似塩基配列」とは、入力された塩基配列と同じ長さで類似する塩基配列であって、前記遺伝子塩基配列に出現する塩基配列である。「類似する」とは、例えば、後で説明するように、入力された塩基配列とのハミング距離が与えられた値になることである。また、「前記遺伝子塩基配列」とは、データベースに格納されている塩基配列である。

【0020】

このような塩基配列検索装置は、例えば、WEBブラウザに入力された塩基配列と類似の程度（例えば、ハミング距離）を受け取り、遺伝子塩基配列を格納したデータベースに対して問い合わせなどを発行して処理を行ない、結果を前記WEBブラウザに返信するサーバ装置として実現が可能である。したがって、本発明に係る塩基配列検索装置の構成要素である各部、各手段は、ハードウェア、ソフトウェア、ハードウェアとソフトウェア（プログラム）の両者、のいずれかによって構成することが可能である。たとえば、これらを実現する一例として、計算機を利用する場合には、CPU、メモリ、バス、インターフェース、周辺装置などから構成されるハードウェアと、これらのハードウェア上にて実行可能なソフトウェアを挙げることができる。

【0021】

（実施形態1：構成）

図4は、本発明の実施形態1に係る塩基配列検索装置の機能ブロック図を例示する。塩基配列検索装置400は、塩基配列入力部401と、ハミング距離入力部402と、特定部403と、割当部404と、選択部405と、置換塩基配列生成部406と、検索部407と、を有する。

【0022】

「塩基配列入力部」401は、所定長を超える長さの塩基配列を入力する。例えば、WEBブラウザから入力された塩基配列を示す情報を受信する。

【0023】

「ハミング距離入力部」402は、入力塩基配列に対して、ハミング距離を入力する。

例えば、WEBブラウザから入力された数値を受信する。ここに、「入力塩基配列」とは、塩基配列入力部401に入力された塩基配列である。また、ハミング距離とは、適合しない塩基への置換の操作を行なう塩基の個数を示す値である。ハミング距離は、2つの同じ長さの塩基配列に対して定義され、対応する塩基が適合しない数であり、1つの塩基配列に対してハミング距離を指定することにより、その塩基配列のうち、ハミング距離の個数の塩基を適合しない塩基に置換して得られる塩基配列の集合が定義できる。

【0024】

ハミング距離の例を挙げる。例えば、図5には、19の塩基からなる塩基配列が上下に並んでいるが、「X」で示される3箇所に対応する塩基が適合していないので、ハミング距離は3となる。図6は、ハミング距離の定義を示す。SとTがn個の塩基からなる塩基配列として、 S_i をi番目の塩基とするときのSとTのハミング距離 $d_H(S, T)$ が定義されている。なお、 $|U|$ を集合とするとき、 $|U|$ により集合の要素の数を表わすとし、 \neq は、両辺の塩基が適合しないことを意味する（例えば、同じでないことを意味する）。したがって、ハミング距離は、負でない整数となる。

【0025】

「特定部」403は、入力塩基配列の部分配列であって、所定長の長さを持ち、異なる2つの部分配列と、その余の部分と、を特定する。2つの部分配列は共通部分があってもよい。また、場合によっては、その余の部分は無くてもよい。

【0026】

図7は、特定部403により特定される2つの部分配列とその余の部分とを例示する。図7(1)においては、第1の部分配列と第2の部分配列が共通部分を持たないように入力塩基配列に存在し、入力塩基配列の両端と中央部分に、その余の部分がある。図7(2)においては、第1の部分配列と第2の部分配列が入力塩基配列の中央に共通部分を有し、入力塩基配列の端の部分にその余の部分がある。図7(3)においては、第1の部分配列が入力塩基配列の左端から伸び、また、第2の部分配列が入力塩基配列の右端から伸び、第1の部分配列と第2の部分配列は、入力塩基配列の中央部分で共通部分を有している。入力塩基配列がある所定長の2倍を超える場合には、図7(4)に例示されるように、入力塩基配列の中央がその余の部分となる。

【0027】

「割当部」404は、特定部403で特定された部分配列とその余の部分とに、ハミング距離入力部402で入力されたハミング距離を分割して割り当てる。ここに、「ハミング距離を分割して割り当てる」とは、ハミング距離を負でない整数に分割して、分割で得られた整数を部分配列とその余の部分に割り振り、割り振られた値の総和は、ハミング距離になるようにすることをいう。このような処理は、プログラムにより容易に実現することができる。例えば、部分配列とその余の部分の個数分のループを入れ子にするプログラムにより実現することができ、全ての割り振りを得ることができる。

【0028】

図8は、図7の(1)から(4)の場合に対応して、割当部404によるハミング距離の割り振りを説明するための図である。図8(1)において、入力塩基配列の左端の部分から（すなわち、その余の部分、第1の部分配列、その余の部分、第2の部分配列、その余の部分）、 m_1 、 m_2 、 m_3 、 m_4 、 m_5 という値が割り振られたとすると、 m_1 、 m_2 、 m_3 、 m_4 、 m_5 の和が入力ハミング距離となる。ここに、「入力ハミング距離」とは、ハミング距離入力部402に入力されたハミング距離である。

【0029】

同様に、図8(2)においても、入力塩基配列の左端の部分から（すなわち、その余の部分、第1の部分配列の左部分、共通部分、第2の部分配列の右部分、その余の部分）、 m_1 、 m_2 、 m_3 、 m_4 、 m_5 が割り振られたとすると、これらの値の和が入力ハミング距離となる。

【0030】

図8(3)においては、入力塩基配列の左端の部分から（すなわち、第1の部分配列の

左部分、共通部分、第2の部分配列の右部分)、 m_1 、 m_2 、 m_3 が割り振られたとすると、これら3つの値の和が入力ハミング距離となる。

【0031】

図8(4)においては、入力塩基配列の左端の部分から(すなわち、第1の部分配列、その余の部分、第2の部分配列)、 m_1 、 m_2 、 m_3 が割り振られたとすると、これら3つの値の和が入力ハミング距離となる。

【0032】

「選択部」405は、特定部403で特定された2つの部分配列のうち、前記割当部で割り当てられたハミング距離で示される個数の塩基を適合しない塩基へ置換する操作を前記部分配列に対して行なって生成される塩基配列である置換塩基配列の総数が大きい方を選択する。この総数は、式： $(\text{適合しない塩基の数}) \times (\text{ハミング距離}) \times (\text{所定長})$ C (ハミング距離)により計算でき、この計算結果に基づいて選択を行なう。しかし、多くの場合には、割当部404で割り当てられたハミング距離の大きい方の部分配列を選択することになる。すなわち、図8(1)の場合には、 m_2 と m_4 とを比較して、例えば、 m_2 の方が大きくなければ、第1の部分配列を選択する。逆に、 m_4 の方が小さければ、第2の部分配列を選択する。ただし、例えば、所定長が4である2つの部分配列に対して、一方にはハミング距離として3が、他方には4が割り当てられたとすると、一方の部分配列の置換塩基配列の総数は、 $(4-1)^3 \times 4 \times C_3 = 108$ であるが、他方の部分配列の置換塩基配列の総数は、 $(4-1)^4 \times 4 \times C_4 = 81$ となり、割り当てられたハミング距離の大きな他方の置換塩基配列の数が少なくなる場合があるので注意が必要である(なお、適合しない塩基を、異なる塩基として計算した)。なお、以下では、説明を簡単にするために、割当部404で割り当てられたハミング距離が大きい場合が、置換塩基配列の総数が大きいと仮定して説明を行なう。

【0033】

同様に図8(2)の場合には、 m_2+m_3 と m_3+m_4 とを比較して、例えば、 m_2+m_3 の方が大きくなければ、第1の部分配列を選択する。逆に、 m_3+m_4 の方が小さければ、第2の部分配列を選択する。

【0034】

また、図8(3)の場合には、 m_1+m_2 と m_2+m_3 とを比較して、例えば、 m_1+m_2 の方が大きくなければ第1の部分配列を選択する。逆に、 m_2+m_3 の方が小さければ、第2の部分配列を選択する。

【0035】

また、図8(4)の場合には、 m_1 と m_3 とを比較して、例えば、 m_1 の方が大きくなければ、第1の部分配列を選択する。逆に、 m_3 の方が小さければ、第2の部分配列を選択する。

【0036】

図9は、入力ハミング距離が3であるとき、図7(4)のように部分配列とその余の部分が特定された場合の割当部404での割り振りと、選択部405による選択を示す。 m_1 、 m_2 、 m_3 の和が3になる組合せは、10通りあるが、選択部405で、例えば $m_1 \leq m_2$ となる選択が行なわれると、6通りに減少し、さらに、 m_1 の値の組合せについてみれば、0と1との2通りに減少する。同様のことが、第2の部分配列と m_3 についても言える。ただし、 $m_1=m_2$ の場合は除かれるので、第2の部分配列が選択される場合の数は、第1の部分配列が選択される場合の数よりも少なくなる。

【0037】

図10は、同じく入力ハミング距離が3であるとき、図7(3)のように部分配列とその余の部分が特定された場合の割当部404での割り振りと、選択部405による選択を示す。 m_1 、 m_2 、 m_3 の和が3になる組合せは、同じく10通りであるが、選択部405で $m_1+m_2 \leq m_2+m_3$ となる選択が行なわれると、6通りの組合せに減少し、第1の部分配列に割り振られた m_1+m_2 の値は、4通りとなる。同じことが第2の部分配列と m_2+m_3 についても言える。ただし、 $m_1+m_2=m_2+m_3$ となる場合が除かれる

ので、第2の部分配列が選択される場合の数は、第1の部分配列が選択される場合の数よりも少なくなる。

【0038】

「置換塩基配列生成部」406は、選択部405により選択された部分配列に対して、割当部404で割り当てられたハミング距離をもつ置換塩基配列を生成する。すなわち、選択部405により選択された部分配列の塩基のうち、割当部404で割り当てられたハミング距離で示される個数の塩基を適合しない塩基に置換することを行ない、置換塩基配列を生成する。例えば、図9の場合には、第1の部分配列について、ハミング距離が0と1となる部分配列が置換塩基配列として生成される。ハミング距離が0であれば、第1の部分配列そのものであり、ハミング距離が1であれば、第1の部分配列の塩基のうちの任意の1つを、適合しない塩基に置き換えて置換塩基配列が生成される。

【0039】

同様に図10の場合には、第1の部分配列について、ハミング距離が0、1、2、3の置換塩基配列として生成される。この場合、入力ハミング距離が3であり、また、ハミング距離が3の置換塩基配列を生成しなければいけないのは、効率が悪いように見える。しかし、3が割り振られたのは、 m_2 であるので、第1の部分配列と第2の部分配列との共通部分に対して、ハミング距離が3となる置換塩基配列を生成すればよい。もし、その共通部分の長さが短ければ、ハミング距離が3となる置換塩基配列の総数は限られたものとなる。このように第1の部分配列と第2の部分配列とが共通部分を持つ場合には、共通部分とそうでない部分とに割り振られたハミング距離を考慮して、共通部分とそうでない部分に個別に置換塩基配列を生成することにより、置換塩基配列の生成の効率を上げることができる。

【0040】

置換塩基配列を生成するプログラムは容易に作成することができ、例えば、ループを入れ子にしたプログラムを作成し、外側のループにより、塩基を適合しない塩基に置換する部分配列の位置を特定し、外側のループにより特定された位置の塩基を適合しない塩基に置換することを内側のループにより行なうようにすればよい。所定長をLとし、塩基が異なるときを適合しないと定義すれば、図9の場合には、 $1 + 3 \text{ } L \text{ } C_1$ 通りの置換塩基配列が生成される。図10の場合には、 $1 + 3 \text{ } L \text{ } C_1 + 3 \text{ } L \text{ } C_2 + 3 \text{ } L \text{ } C_3$ 通りの置換塩基配列が生成されるが、この生成に必要な計算量は、一般にLは入力塩基配列の長さの値よりも小さいので、入力塩基配列とハミング距離が3となる塩基配列の全てを求める計算量より小さい。

【0041】

「検索部」407は、置換塩基配列生成部で生成された置換塩基配列をキーとして前記索引を用いて検索を行なう。多くの場合、索引はハッシュの手法を用いて実現されている。「前記索引」とは、所定長の塩基配列の、遺伝子配列を格納したデータベースにおける出現を検索するための索引である。このような索引による検索により、一般には、置換塩基配列が出現する位置情報（例えば、置換塩基配列の端の塩基が、DNAの5'端から何番目の位置の塩基になるかを示す情報）が得られる。

【0042】

もし、塩基配列検索装置が遺伝子塩基配列を格納したデータベースを備えていれば、検索部407は、そのデータベースに対して問い合わせを行なう。また、他のサーバであっても、そのようなデータベースを備えているサーバがあれば、検索部407はそのサーバに問い合わせを送信して、結果を受信するようになっていてもよい。

【0043】

（実施形態1：処理の流れ）

図11は、本実施形態に係る塩基配列検索装置の処理の流れ図を例示する。ステップS1101において、塩基配列入力部401などにより、塩基配列を入力する（塩基配列入力ステップ）。ステップS1102において、ハミング距離入力部402などにより、ハミング距離を入力する（ハミング距離入力ステップ）。ステップS1103において、特

定部 4 0 3 などにより、2つの部分配列とその余の部分とを特定する（特定ステップ）。ステップ S 1 1 0 4 において、割当部 4 0 4 などにより、入力されたハミング距離を分割して割り当てる（割当ステップ）。ステップ S 1 1 0 5 において、選択部 4 0 5 などにより、割当ステップで割り当てられたハミング距離を有する置換塩基配列の総数の大きい方の部分配列を選択する（選択ステップ）。ステップ S 1 1 0 6 において、置換塩基配列生成部 4 0 6 などにより、置換塩基配列を生成する（置換塩基配列生成ステップ）。ステップ S 1 1 0 7 において、検索部 4 0 7 などにより、検索を行なう（検索ステップ）。

【0044】

したがって、塩基配列検索装置は、塩基配列入力ステップ、ハミング距離入力ステップ、特定ステップ、割当ステップ、選択ステップ、置換塩基配列生成ステップ、検索ステップを含む塩基配列検索方法を使用するための装置とみなすこともできる。

【0045】

なお、図 1 1 に例示された流れ図は一例であり、ステップ S 1 1 0 1 で入力された塩基配列の一つについて、ステップ S 1 1 0 2 で入力されるべきハミング距離を 0、1、2、3、4 などと変化させながら、その他のステップを繰り返し実行してもよい。また、ステップ S 1 1 0 1 を行なった後でステップ S 1 1 0 3 を行ない、ステップ S 1 1 0 2 を実行して、その他のステップを実行するようになっていてもよい。

【0046】

（実施形態 1：主な効果）

本実施形態により、検索のために必要となる計算量を小さくすることができ、また、ハミング距離が所定の値となり似た塩基配列を漏れなく検索することができる。

【0047】

（実施形態 2：主に請求項 2 について説明する）

【0048】

図 1 2 は、本発明の実施形態 2 に係る塩基配列検索装置の機能ブロック図を例示する。塩基配列検索装置 1 2 0 0 は、塩基配列入力部 4 0 1 と、ハミング距離入力部 4 0 2 と、特定部 4 0 3 と、割当部 4 0 4 と、選択部 4 0 5 と、置換塩基配列生成部 4 0 6 と、検索部 4 0 7 と、を有し、特定部 4 0 3 は、第一特定手段 1 2 0 1 を有している。したがって、本実施形態に係る塩基配列検索装置は、実施形態 1 に係る塩基配列検索装置の特定部が第一特定手段を有した構成となっている。

【0049】

「第一特定手段」1 2 0 1 は、塩基配列入力部で入力された塩基配列の塩基数が前記所定長の 2 倍以下または 2 倍未満であれば、前記 2 つの部分配列のうち一方の部分配列の端を前記入力塩基配列の他方の端と一致させ、その余の部分が生じず特定されないことにする。その余の部分が生じず特定されないことにより、割当部では、その余の部分にハミング距離を割り当てることはしないこととなる。

【0050】

すなわち、第一特定手段は、図 7（3）のように第 1 の部分配列と第 2 の部分配列とを特定する。したがって、このような場合は実施形態 1 について既に説明されているので、以後の説明は省略する。

【0051】

（実施形態 3：主に請求項 3 について説明する）

【0052】

図 1 3 は、本発明の実施形態 3 に係る塩基配列検索装置の機能ブロック図を例示する。塩基配列検索装置 1 3 0 0 は、塩基配列入力部 4 0 1 と、ハミング距離入力部 4 0 2 と、特定部 4 0 3 と、割当部 4 0 4 と、選択部 4 0 5 と、置換塩基配列生成部 4 0 6 と、検索部 4 0 7 と、を有し、特定部 4 0 3 は、第二特定手段 1 3 0 1 を有している。また、特定部 4 0 3 は、実施形態 2 で説明した第一特定手段を有していてもよい。したがって、本実施形態に係る塩基配列検索装置は、実施形態 1 または 2 に係る塩基配列検索装置の特定部が第二特定手段を有した構成となっている。

【0053】

「第二特定手段」1301は、塩基配列入力部で入力された塩基配列の塩基数が前記所定長の2倍より大であれば、前記2つの部分配列が重ならないことにして、前記2つの部分配列を特定する。この場合、その余の部分が一つになるようにしてもよいし、2つになるようにしてもよい。例えば、2つの部分配列が入力塩基配列の左右の端に配置されるように特定したり、2つの部分配列が接続されるように入力塩基配列を特定したりする。

【0054】

すなわち、第二特定手段は、図7(4)のように第1の部分配列と第2の部分配列とを特定する。したがって、このような場合は実施形態1について既に説明されているので、以後の説明は省略する。

【0055】

(実施形態4：主に請求項4について説明する)

【0056】

本発明の実施形態4として、検索部での検索結果に基づいて、類似塩基配列の候補を取得して、入力塩基配列とのハミング距離を判定する塩基配列検索装置について説明する。

【0057】

(実施形態4：構成)

図14は、本発明の実施形態4に係る塩基配列検索装置の機能ブロック図を例示する。塩基配列検索装置1400は、塩基配列入力部401と、ハミング距離入力部402と、特定部403と、割当部404と、選択部405と、置換塩基配列生成部406と、検索部407と、類似候補塩基配列取得部1401と、判定部1402と、を有している。また、特定部403は、実施形態2で説明した第一特定手段と実施形態3で説明した第二特定手段とのいずれか一方または両方を有していてもよい。したがって、本実施形態に係る塩基配列検索装置は、実施形態1から3のいずれか一の塩基配列検索装置が類似候補塩基配列取得部1401と、判定部1402と、を有している構成となっている。

【0058】

「類似候補塩基配列取得部」1401は、検索部407での検索結果に基づいて、類似候補塩基配列を取得する。「類似候補塩基配列」とは、置換塩基配列を含んで遺伝子塩基配列に現れる塩基配列である。より具体的に説明すると、例えば、第1の部分配列の置換塩基配列により検索が行なわれ、置換塩基配列の端の塩基の位置が判明したとすると、第1の部分配列が入力塩基配列に占める位置関係を勘案して得られる塩基配列で、入力塩基配列と同じ長さの遺伝子塩基配列を取得する。すなわち、検索で得られる位置が第1の部分配列の左端の塩基の位置であるとする、第1の部分配列の左側のその余の部分の長さ(もし、そのようなその余の部分があれば0とする)だけ左の位置から、入力塩基配列と同じ長さの遺伝子塩基配列を取得する。第2の部分配列の置換塩基配列について検索が行なわれた場合も同様に、第2の部分配列の右側のその余の部分の長さだけ右の位置から左に向かって入力塩基配列と同じ長さの遺伝子塩基配列を取得する。この取得は、データベースを検索することにより行なわれる。もし、塩基配列検索装置が、そのようなデータベースを備えていれば、そのデータベースから取得を行ない、別のサーバにそのようなデータベースが備えられていればそのサーバに取得の要求を送信して、類似候補塩基配列を得る。

【0059】

「判定部」1402は、類似候補塩基配列取得部で取得された類似候補塩基配列と、前記入力塩基配列と、のハミング距離が、ハミング距離入力部402に入力されたハミング距離と等しいかどうかを判定する。この判定は、入力塩基配列と類似候補塩基配列との端の塩基から順に比較を行なうことにより行なうことができる。

【0060】

本実施形態に係る塩基配列検索装置の処理の流れ図は、図11に例示された流れ図のステップS1107の後に、類似候補塩基配列を取得するステップと、類似候補塩基配列と入力塩基配列とのハミング距離が、入力ハミング距離に等しいかどうかを判定するステッ

ブと、を実行する流れ図となる。

【0061】

（実施形態4：主な効果）

本実施形態によれば、入力塩基配列に類似する塩基配列を取得することができ、例えば、s i R N Aにより不活性化する目的の遺伝子以外に不活性化される遺伝子の情報を得ることが可能となる。

【0062】

（実施形態5：主に請求項5について説明する）

【0063】

本発明の実施形態5として、不適合となる塩基の組合せを指定することができる塩基配列検索装置について説明する。

【0064】

（実施形態5：構成）

図15は、本発明の実施形態5に係る塩基配列検索装置の機能ブロック図を例示する。塩基配列検索装置1500は、塩基配列入力部401と、ハミング距離入力部402と、特定部403と、割当部404と、選択部405と、置換塩基配列生成部406と、検索部407と、類似候補塩基配列取得部1401と、判定部1402と、不適合塩基組入力部1501と、を有している。したがって、本実施形態に係る塩基配列検索装置は、実施形態4に係る塩基配列検索装置が不適合塩基組入力部1501を有している構成となっている。

【0065】

「不適合塩基組入力部」1501は、適合しない塩基の組を指定する。例えば適合しないと判断すべき塩基のペアを示すテキスト情報を入力する。あるいは、適合と判断すべき塩基のペア（例えば、GとU）を入力することにより、間接的に適合しないと判断すべき塩基の組が指定されるようになっていてもよい。

【0066】

本実施形態においては、不適合塩基組入力部1501に入力された塩基の組に基づいて検索部で検索が行なわれ、また、ハミング距離が求められる。例えば、不適合塩基組入力部1501により入力された塩基の組に基づいて、置換塩基配列が置換塩基配列生成部406で生成され、検索部407では検索のためのデータベースが選択され、判定部1402でハミング距離が求められる。

【0067】

（実施形態5：主な効果）

本実施形態によれば、例えば、GとUのように弱いながらも結合する可能性のある塩基の組合せを考慮することができ、より正確な塩基配列の設計を行なうことが可能となる。

【0068】

（実施形態6：主に請求項6について説明する）

【0069】

本発明の実施形態6として、入力塩基配列と類似塩基配列との塩基の適合の分布を指定することができる塩基配列検索装置について説明する。

【0070】

（実施形態6：構成）

図16は、本発明の実施形態6に係る塩基配列検索装置の機能ブロック図を例示する。塩基配列検索装置1600は、塩基配列入力部401と、ハミング距離入力部402と、特定部403と、割当部404と、選択部405と、置換塩基配列生成部406と、検索部407と、類似候補塩基配列取得部1401と、判定部1402と、適合分布入力部1601と、を有しており、判定部1402は、判定手段1602を有している。また、塩基配列検索装置1600は、実施形態5にて説明した不適合塩基組入力部を有しているもよい。したがって、本実施形態に係る塩基配列検索装置は、実施形態4または5に係る塩基配列検索装置が、適合分布入力部1601を有し、判定部1402は、判定手段1602

2を有している構成となっている。

【0071】

「適合分布入力部」1601は、塩基配列入力部401に入力された塩基配列と類似塩基配列との対応する塩基の適合の分布を表わす分布情報を入力する。分布情報の例としては、5'端側の方に塩基の不適合の発生が少ない、あるいは、多い、塩基の不適合がほぼ等間隔で発生していることを示す情報がある。分布情報は、例えば、塩基の適合の分布を判定するプログラムであってもよい。あるいは、あらかじめ塩基の適合の分布の類型をいくつか決めておき、それらを選択するための情報であってもよい。

【0072】

「分布判定手段」1602は、適合分布入力部1602で入力された分布情報が満たされているかどうかを判定する。

【0073】

判定部1402は、例えば、類似塩基配列とともに、分布判定手段での判定の結果を表示するようになっていてもよい。

【0074】

（実施形態6：主な効果）

本実施形態により、より正確な塩基配列の設計を行なうことが可能となる。

【0075】

（実施形態7：主に請求項7について説明する）

【0076】

本発明の実施形態7に係る塩基配列検索装置は、実施形態6に係る塩基配列検索装置において、適合分布入力部1601で入力される分布情報を、塩基配列と類似塩基配列との対応する塩基が連続して適合する長さの下限としたものである。

【0077】

2つの塩基配列において、対応する塩基に不適合となるものがあっても、対応する塩基が連続して適合していると、結合（ハイブリダイズ）してしまう場合がある。本実施形態においては、塩基が連続して適合する長さの下限を指定することにより、結合してしまう可能性のある類似塩基配列を検出するようにしたものである。

【0078】

（実施形態8：主に請求項8について説明する）

【0079】

本発明の実施形態8は、実施形態1から7のいずれかの実施形態において、塩基配列入力部に入力される塩基配列の長さを15から25とし、所定長を11から14とした実施形態である。

【0080】

塩基配列入力部に入力される塩基配列の長さを15から25とすることにより、本実施形態に係る塩基配列検索装置をs i R N Aの設計に適したものとすることができる。また、発明者がベンチマークテストに用いたデータベースでは、入力塩基配列の長さが19または20のときには、所定長を11から14とした場合が、最も高速に検索が行なえた。これは、所定長が小さいと、類似候補塩基配列の候補の数が多くなり、一方、所定長を大きくすると、置換塩基配列生成部での置換塩基配列の生成に計算量が必要となるとともに、索引を構成するハッシュテーブルに対して問い合わせを行なった際のミスヒットが増加する、すなわち、もともとのデータベース中に存在しない配列を問い合わせすることになる場合が増え、計算量が増加するためであり、その中間点が、所定長が11から14である場合と考えられる。

【0081】

（実施形態9：主に請求項10、11について説明する）

【0082】

以上、データベースに格納された遺伝子塩基配列に対する検索について述べたが、本発明の技術は、遺伝子塩基配列に限らず、一般の文字列検索などに応用することができる。

すなわち、遺伝子塩基配列は、4つの塩基が一次元に配列したものであるもので、それぞれの塩基を、文字列を構成するアルファベットとみなすことにより、遺伝子塩基配列を文字列とみなすことができる。また、上記の説明から判明するように、塩基の数が4である点は、本発明の技術を一般の文字列に対して適用する制限とはならない。

【0083】

したがって、本発明の技術により、データベースに蓄積された文字列から、入力された文字列に類似する文字列を検索することが可能となる。ここに「類似する」とは、入力された文字列から所定のハミング距離もつ文字列を意味する。

【0084】

したがって、次の文字列検索装置が提供される。すなわち、アルファベットが一次元に配列した文字列を格納したデータベースを検索するための索引であり、所定の長さである所定長の文字列が前記データベースに格納された文字列の中に出現する位置を検索するための索引、を用いて、入力される文字列と同じ長さで類似する文字列であり前記データベースに格納された文字列に出現する文字列である類似文字列を検索するための文字列検索装置であって、前記所定長を超える長さの文字列を入力する文字列入力部と、前記文字列入力部に入力された文字列である入力文字列に対して、適合しないアルファベットへの置換の操作を行なうアルファベットの個数を示すハミング距離を入力するハミング距離入力部と、前記入力文字列の部分文字列であって、前記所定長の長さを持ち異なる2つの部分文字列と、その余の部分と、を特定する特定部と、前記特定部で特定された部分文字列とその余の部分とに、前記ハミング距離入力部で入力されたハミング距離を分割して割り当てる割当部と、前記特定部で特定された2つの部分文字列のうち、前記割当部で割り当てられたハミング距離で示される個数のアルファベットを適合しないアルファベットへ置換する操作を前記部分文字列に対して行なって生成される文字列である置換文字列の総数が大きい方を選択する選択部と、前記選択部により選択された部分文字列に対して、前記割当部で割り当てられたハミング距離をもつ置換文字列を生成する置換文字列生成部と、前記置換文字列生成部で生成された置換文字列を検索キーとして前記索引を用いて検索を行なう検索部と、を有する文字列検索装置を提供することが可能となる。

【0085】

また、文字列のアルファベットをペプチドとすることにより、本発明の技術をペプチド配列の類似検索、すなわち、入力されたペプチド配列に類似のペプチドを検索することにも使用することができる。

【産業上の利用可能性】

【0086】

本発明に係る塩基配列検索装置及び塩基配列検索方法は、検索のために必要となる計算量を小さくすることができ、また、ハミング距離が所定の値となり似た塩基配列の存在を見落とすことも無いので、塩基配列などの設計に有用である。

【図面の簡単な説明】

【0087】

【図1】RNA干渉の過程の概略を示す図

【図2】マイクロアレイを用いた遺伝子解析や遺伝子診断などの過程を説明するための図

【図3】BLASTで見落とされる可能性のある塩基配列を説明するための図

【図4】本発明の実施形態1に係る塩基配列検索装置の機能ブロック図

【図5】ハミング距離が3の塩基配列の一例図

【図6】ハミング距離の定義を示す図

【図7】特定される2つの部分配列とその余の部分との一例図

【図8】ハミング距離の割り振りを説明するための図

【図9】割当部によるハミング距離の割り振りと選択部による選択を説明するための図

【図10】割当部によるハミング距離の割り振りと選択部による選択を説明するための図

の図

【図 1 1】 本発明の実施形態 1 に係る塩基配列検索装置の処理の流れ図

【図 1 2】 本発明の実施形態 2 に係る塩基配列検索装置の機能ブロック図

【図 1 3】 本発明の実施形態 3 に係る塩基配列検索装置の機能ブロック図

【図 1 4】 本発明の実施形態 4 に係る塩基配列検索装置の機能ブロック図

【図 1 5】 本発明の実施形態 5 に係る塩基配列検索装置の機能ブロック図

【図 1 6】 本発明の実施形態 6 に係る塩基配列検索装置の機能ブロック図

【符号の説明】

【 0 0 8 8 】

4 0 0 塩基配列検索装置

4 0 1 塩基配列入力部

4 0 2 ハミング距離入力部

4 0 3 特定部

4 0 4 割当部

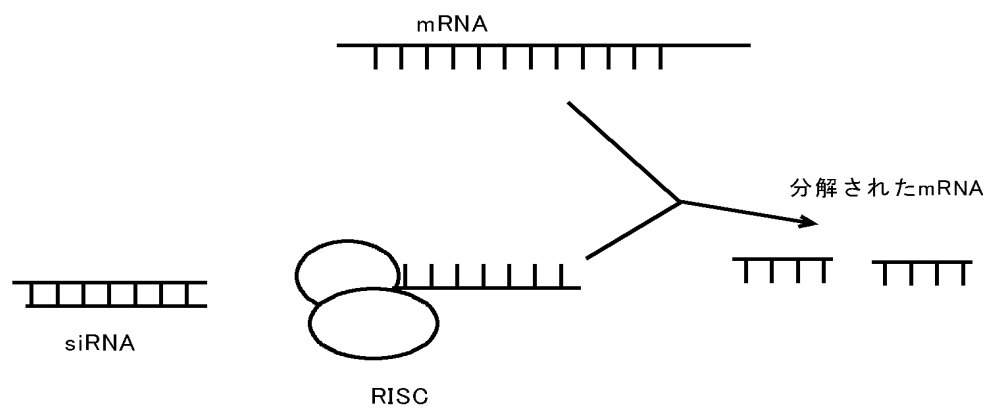
4 0 5 選択部

4 0 6 置換塩基配列生成部

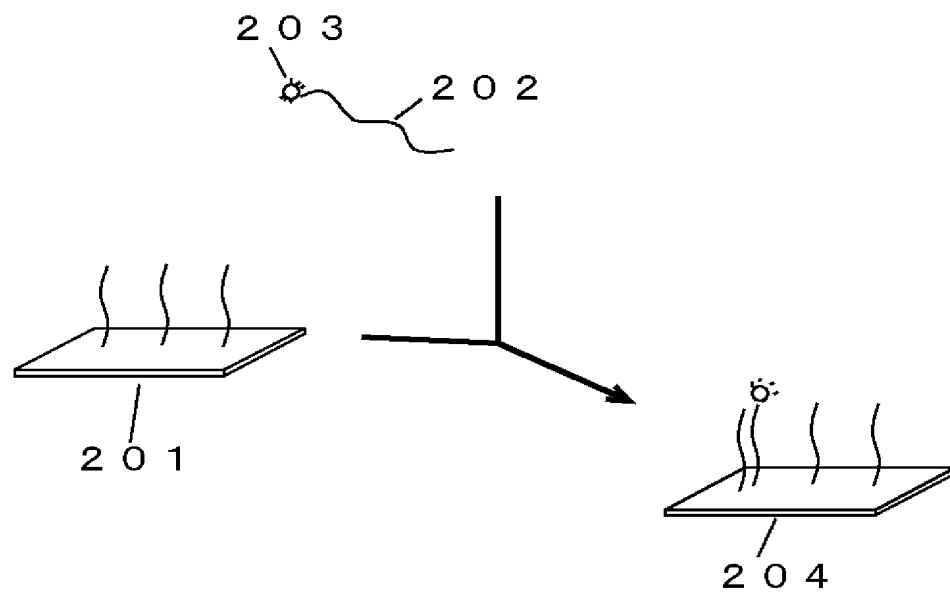
4 0 7 検索部

【書類名】 図面

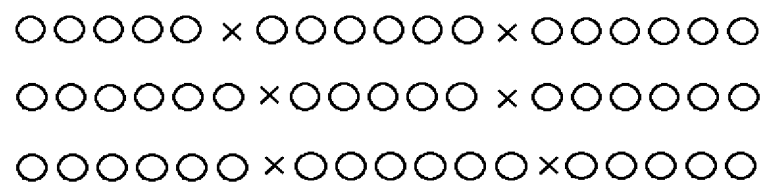
【図 1】



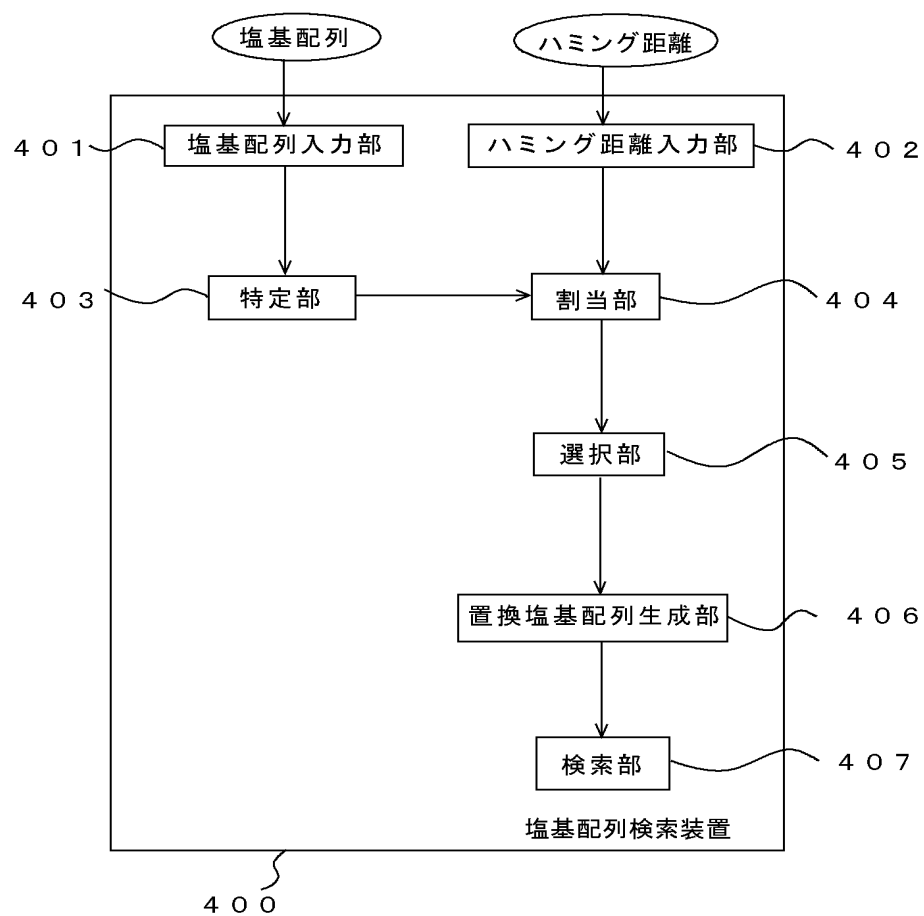
【図 2】



【図 3】



【図 4】



【図 5】

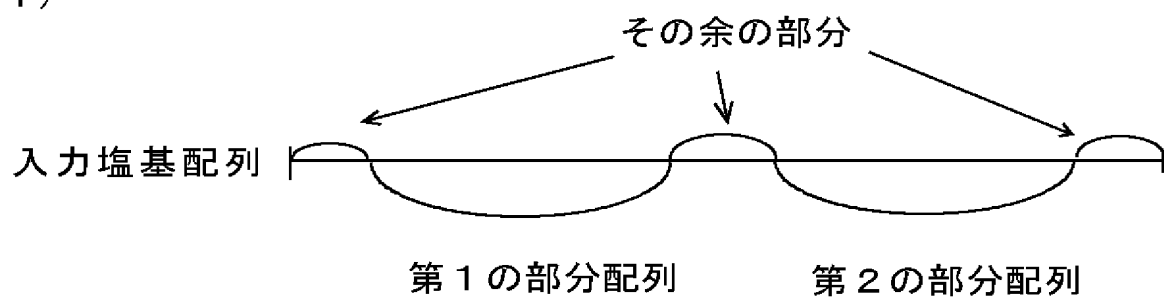
A	C	C	G	T	A	G	T	A	T	A	T	G	C	T	G	C	T	G
				X						X					X			
A	C	C	G	C	A	G	T	A	T	C	T	G	C	T	T	C	T	G

【図 6】

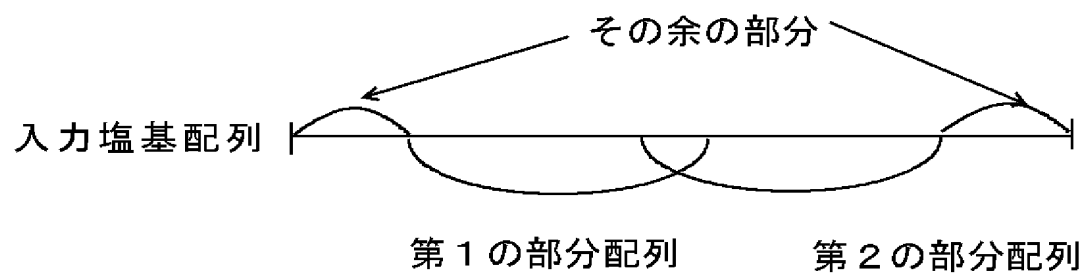
$$d_H(S, T) = \left| \{ i \mid S_i \neq T_i, i=1, 2, \dots, n \} \right|$$

【図 7】

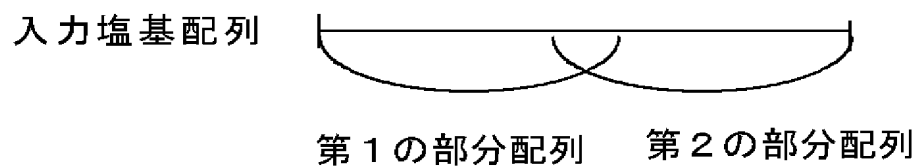
(1)



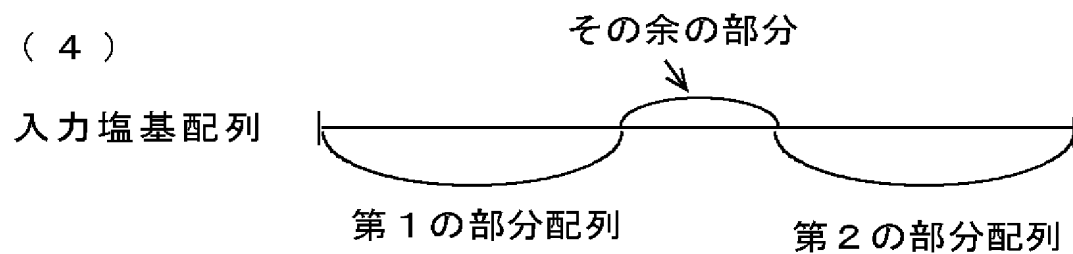
(2)



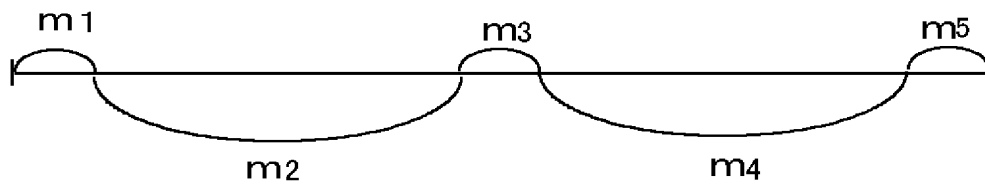
(3)



(4)

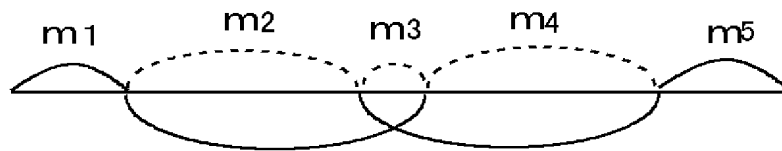


(1)



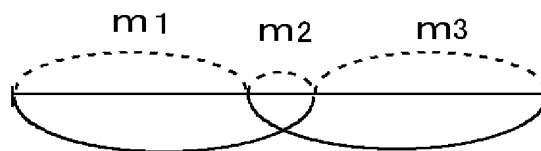
$$m_1 + m_2 + m_3 + m_4 + m_5 = \text{入力ハミング距離}$$

(2)



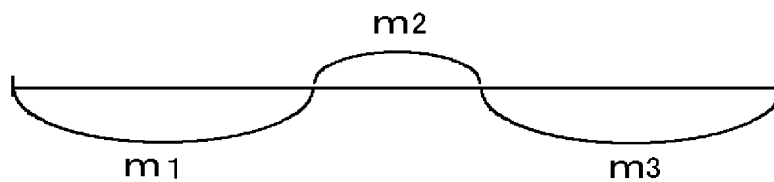
$$m_1 + m_2 + m_3 + m_4 + m_5 = \text{入力ハミング距離}$$

(3)

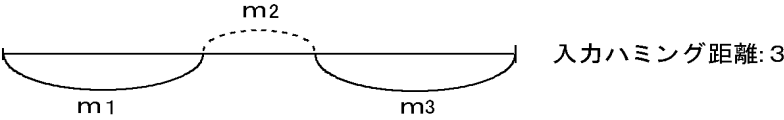


$$m_1 + m_2 + m_3 = \text{入力ハミング距離}$$

(4)



$$m_1 + m_2 + m_3 = \text{入力ハミング距離}$$



m1	m2	m3
0	0	3
0	3	0
3	0	0
0	1	2
1	0	2
0	2	1
1	2	0
2	0	1
2	1	0
1	1	1



$m1 \leq m2$

m1	m2	m3
0	0	3
0	3	0
0	1	2
1	0	2
0	2	1
1	1	1



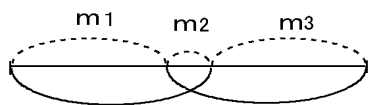
m1
0
1

$m1 > m2$

m1	m2	m3
3	0	0
1	2	0
2	0	1
2	1	0



m2
0
1



入力ハミング距離: 3

m1	m2	m3
0	0	3
0	3	0
3	0	0
0	1	2
1	0	2
0	2	1
1	2	0
2	0	1
2	1	0
1	1	1



$m1 + m2 \leq m2 + m3$

$m1 + m2 > m2 + m3$

m1	m2	m3
0	0	3
0	3	0
0	1	2
1	0	2
0	2	1
1	1	1

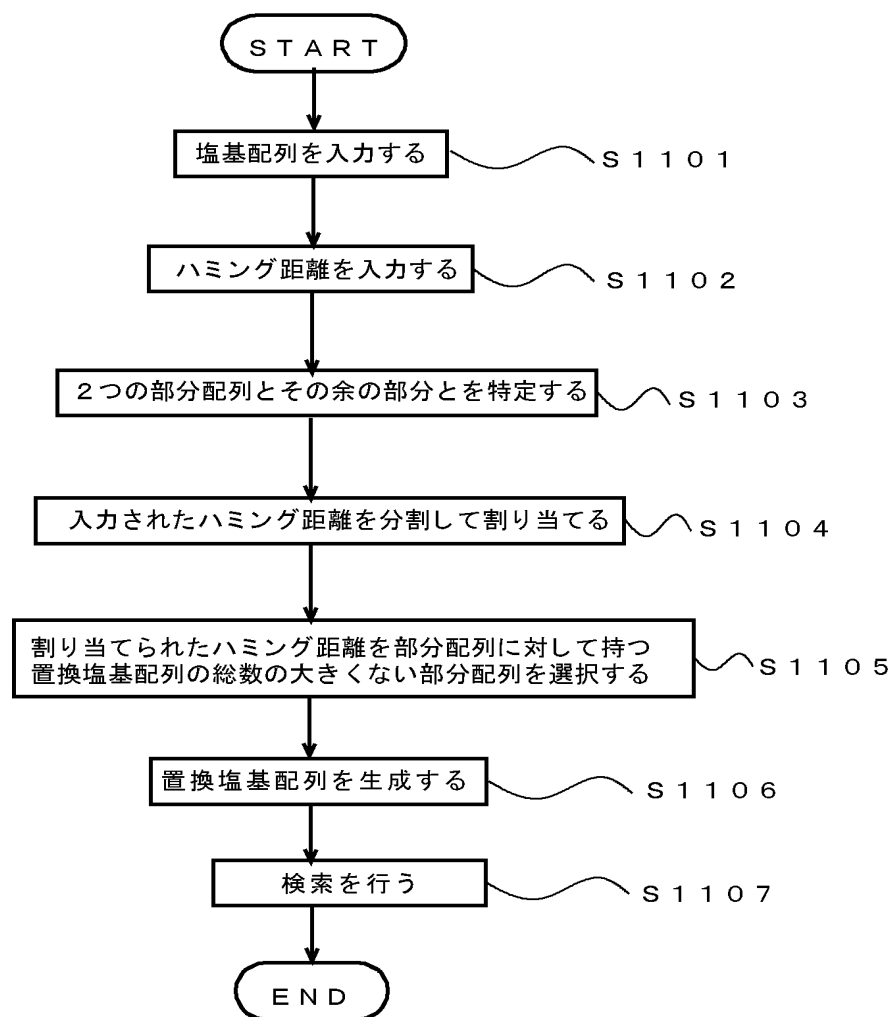


m1 + m2
0
1
2
3

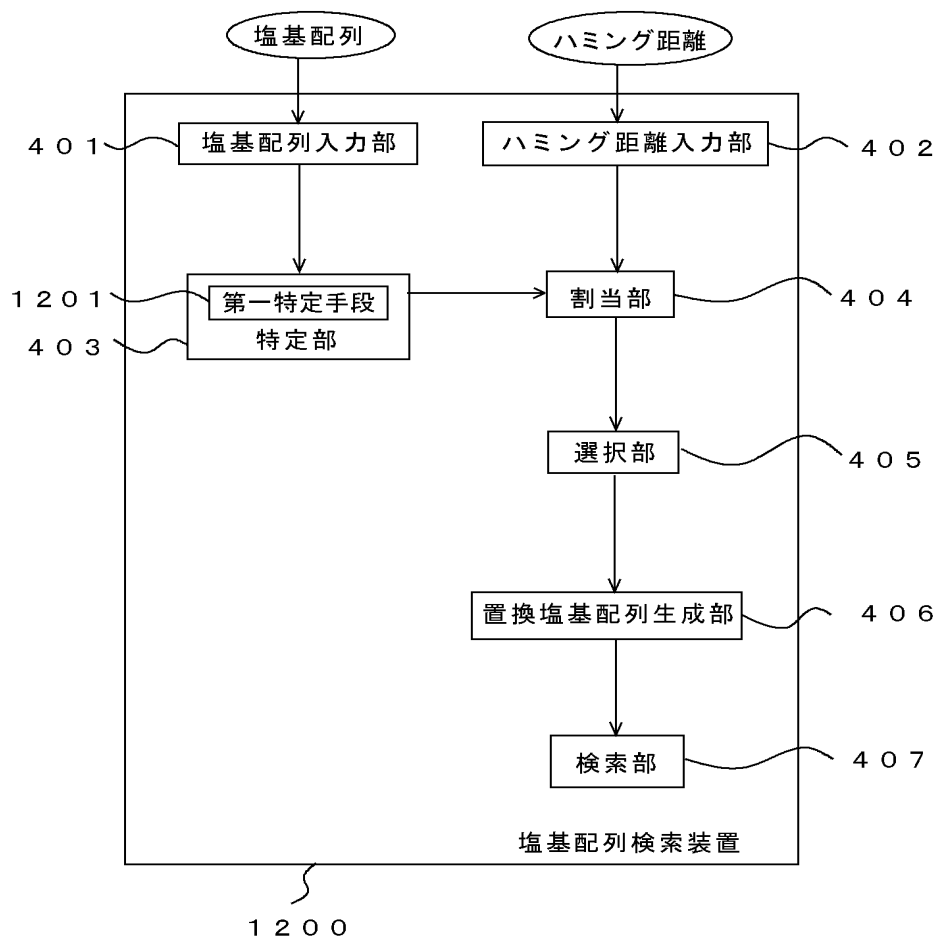
m1	m2	m3
3	0	0
1	2	0
2	0	1
2	1	0



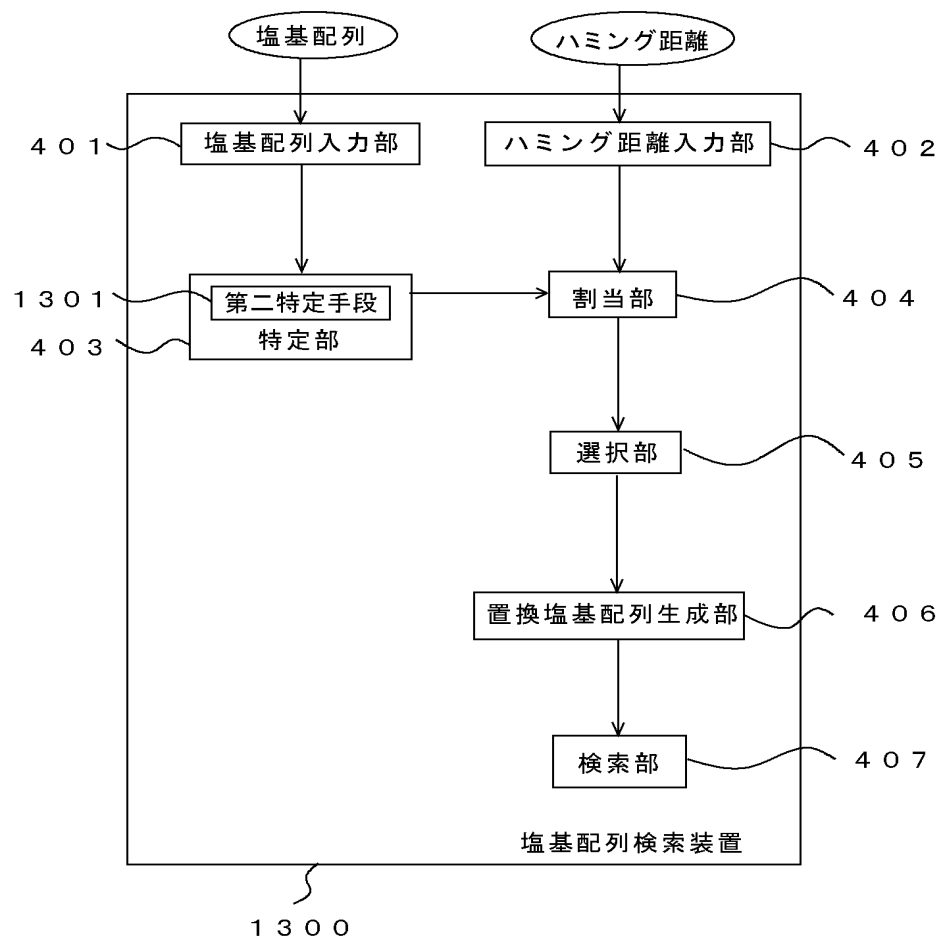
m2 + m3
0
1



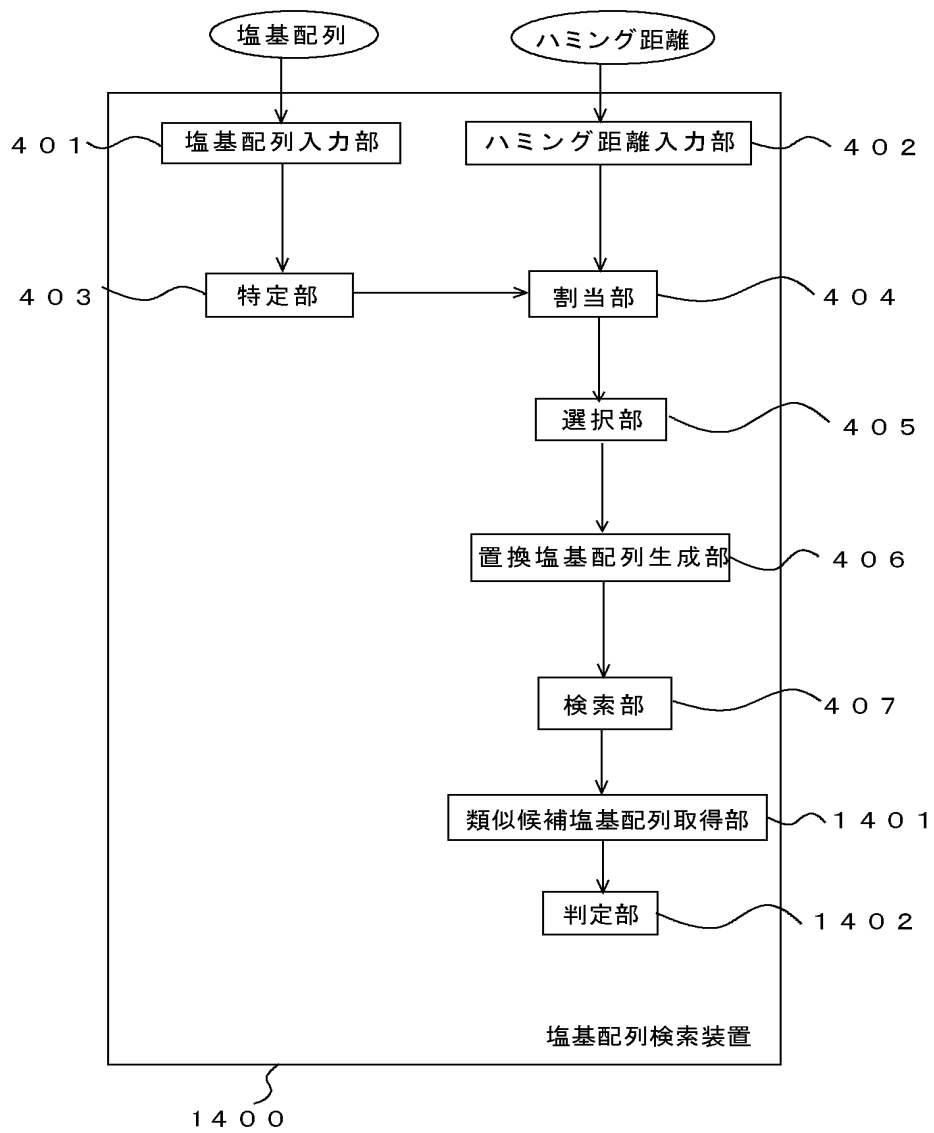
【図 1 2】



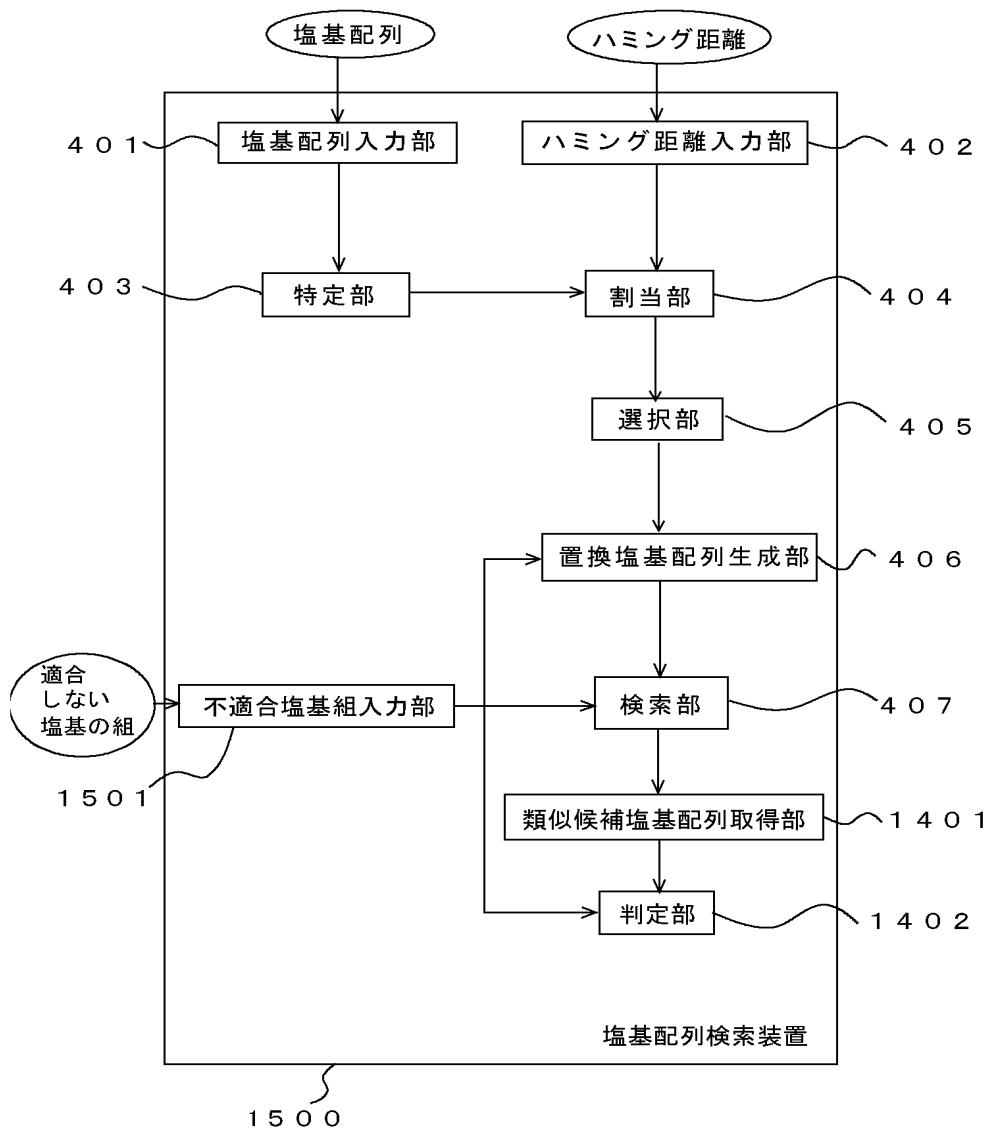
【図 1 3】



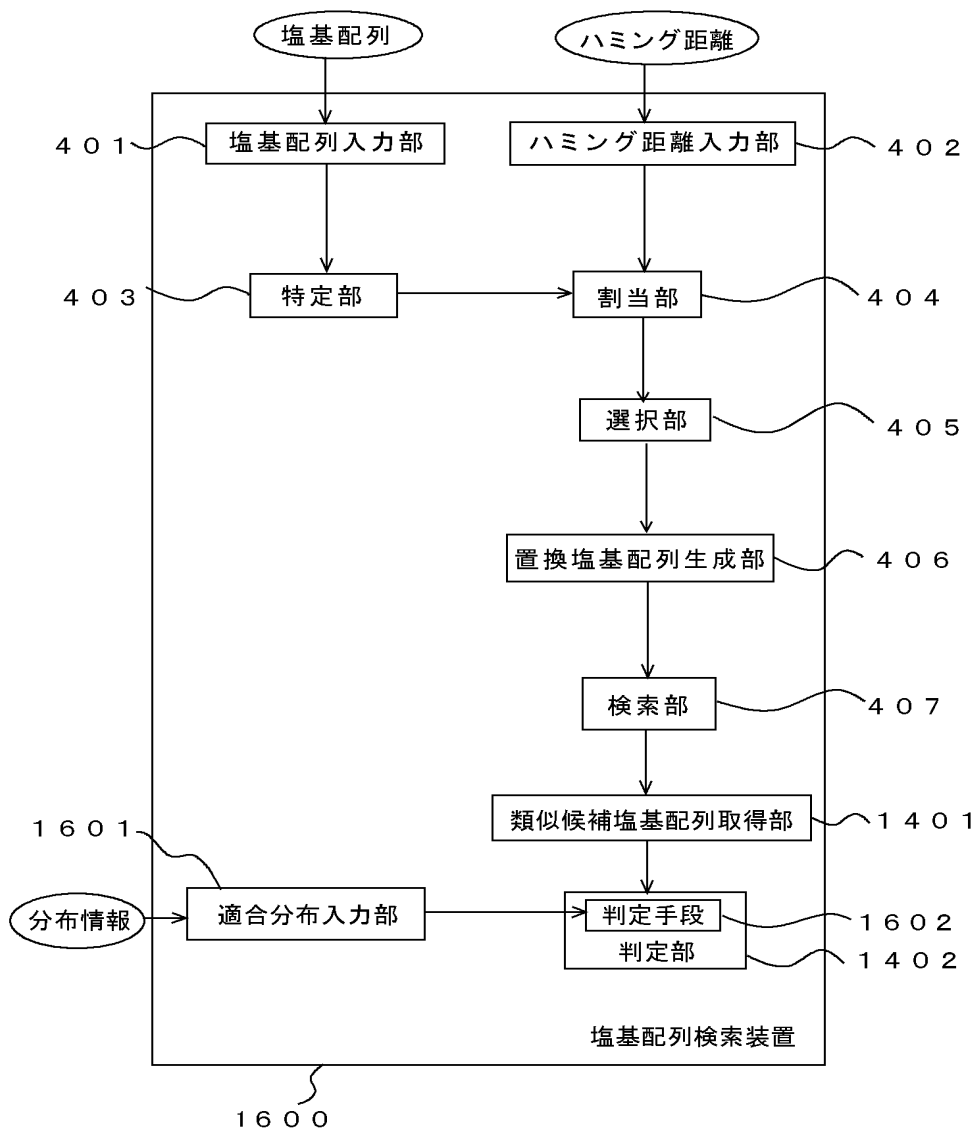
【図 1 4】



【図 15】



【図 16】



【書類名】 要約書

【要約】

【課題】 s i R N A などの塩基配列を設計する場合に、類似する塩基配列を含む遺伝子を漏れなく高速に検索する装置及び方法などを提供する。

【解決手段】 入力された塩基配列から所定の長さの二つの部分配列とその余の部分とを特定して、対応する塩基が適合しない数であるハミング距離を、それらの部分配列とその余の部分とに分割して割り当てて、2つの部分配列のうち、割り当てられた数が大きいほうを選択して、検索を行なうようにする。

【選択図】 図 8

【書類名】	出願人名義変更届
【整理番号】	BI0TT-0002
【あて先】	特許庁長官殿
【事件の表示】	
【出願番号】	特願2004-108456
【承継人】	
【住所又は居所】	東京都文京区本郷三丁目3 2 番 6 - 7 0 3
【氏名又は名称】	株式会社バイオシンクタンク
【代表者】	名取 幸和
【承継人代理人】	
【識別番号】	100109553
【弁理士】	
【氏名又は名称】	工藤 一郎
【手数料の表示】	
【予納台帳番号】	100322
【納付金額】	4,200円
【提出物件の目録】	
【物件名】	承継人であることを証する書面 1
【援用の表示】	おって手続補足書で提出する
【物件名】	代理権を証明する書面（委任状） 1
【援用の表示】	おって手続補足書で提出する

出願人履歴

5 0 4 1 1 9 9 7 4

20040326

新規登録

東京都練馬区田柄 1 丁目 1 9 － 7

森下 真一

5 0 2 0 0 1 4 3 3

20011227

新規登録

東京都文京区本郷 5 － 1 4 － 8 － 3 0 2

山田 智之

5 0 4 1 2 0 0 0 6

20040326

新規登録

神奈川県横浜市西区みなとみらい四丁目 1 0 番 1 － E 1 7 0 6 号

名取 幸和

5 0 4 4 2 0 6 2 9

20041112

新規登録

東京都文京区本郷三丁目 3 2 番地 6 － 7 0 3 号

株式会社 バイオシンクタンク

5 0 4 4 2 0 6 2 9

20041213

住所変更

東京都文京区本郷三丁目3 2 番6 － 7 0 3 号

株式会社 バイオシンクタンク